

ON THE BIOMECHANICAL CONTROL VARIABLES OF THE TONGUE DURING SPEECH MOVEMENTS*

Rafael Laboissière¹

Vittorio Sanguineti²

Yohan Payan¹

E-mail: rafael@icp.grenet.fr W³: http://cristal.icp.grenet.fr:8080

¹ Institut de la Communication Parlée
INPG, 46 av. Félix Viallet F-38031 Grenoble CEDEX 1 France

² Dipartimento di Informatica Sistemica e Telematica
University of Genova, Via Opera Pia, 13 I-16145 Genova Italy

ABSTRACT

A model of the tongue musculature is introduced which is based on the Equilibrium Point Hypothesis (λ model). The model is fitted to mid-sagittal cineradiographic data in order to extract muscles synergies through a Principal Component Analysis. An assessment of the effects of these commands on tongue shape is done, showing how the degrees of freedom of tongue profile relate to the biomechanical substrate.

1. INTRODUCTION

1.1. Background

The tongue is one of the most important articulators involved in speech production. Indeed, it is the main contributor to the shaping of the oral cavities and consequently to the acoustical outcome. Among the large amount of previous studies concerning tongue control, many of them focused on statistical descriptions of its shape during speech movements (e.g. Maeda 1988). The tongue musculature system is highly complicated and those studies do not provide any clue to understand both (i) how the degrees of freedom of tongue shape relates to the biomechanical substrate, and (ii) how the Central Nervous System (CNS) organizes muscle commands in order to produce movements in those degrees of freedom.

These same issues were recently addressed in the case of the jaw/hyoid system (Laboissière, Ostry, and Feldman 1995), for which the mechanical degrees of freedom are easily identifiable (translation and rotation of the bony parts). Unlike the jaw and the hyoid bone, the tongue is a non-rigid body and the number of mechanical degrees of freedom is much higher than the number of functional degrees of freedom obtained from the statistical analysis of tongue shape. Our main goal in the present work is to understand how those functional degrees of freedom are expressed as specific muscle synergies.

*This work was supported by the European Community ESPRIT/BR grant # 6975 SPEECH MAPS and by GALILEO France-Italy project # 94290.

1.2. The λ model

We propose to address those questions in the framework of the motor control scheme known as λ model Feldman (1986), a computationally powerful and biologically plausible version of the Equilibrium-Point (EP) hypothesis of motor control.

The central point in the λ model is that the CNS does not directly specify the degree of muscle activation, but indeed establishes a threshold muscle length (λ), at which motoneuron recruitment begins; this hypothesis is based on empirical *in vivo* observations suggesting that, in intact muscles, the relationship between length and force is described by a family of length-tension curves—the so-called *Invariant Characteristics* (ICs)—each characterized by a different rest length, λ .

Each λ command can be written (see Fig. 1) as the difference between a *reciprocal* (R) and a *coactivation* (C) command, respectively responsible for (almost) independently specifying a desired muscle length, $l = l^*$ (in fact $R = l^*$), and a desired muscle force, $f_M = f_M^*$ (which is proportional to C); therefore, by changing λ the CNS may produce either motion to a new equilibrium position or increase in muscle activation and force.

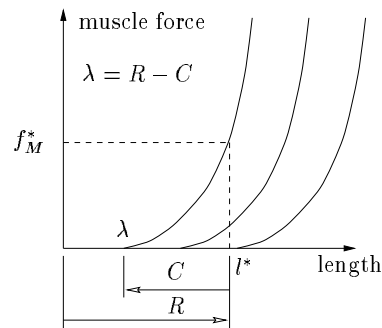


Figure 1: Invariant Characteristic. Reciprocal (R) and Coactivation (C) commands corresponding to $l = l^*$ and $f_M = f_M^*$

From a computational point of view, the most distinctive feature of the λ model is the key role assigned to geometric/kinematic variables in the central control of movements; a desired posture of the body is specified by setting each λ to the corresponding muscle length, l . This is known as the *principle of anatomical correspondence* and provides a natural solution of the excess of degrees of freedom problem: given a desired equilibrium configuration, \vec{u} , for each muscle the corresponding command is $\lambda = l = \mathcal{L}(\vec{u})$.

In other words, the λ commands (one for each muscle) are entirely specified by a smaller number of variables, \vec{u} , which define a desired equilibrium posture and are referred by Feldman (1995) as *control variables* (CVs).

Moreover, the model predicts that specific knowledge of muscle behavior (with their complex dynamics) is not required at central level, therefore the problem of specifying body postures in terms of control variables is greatly simplified, and made independent from the related problem of specifying, through C , the postural behavior.

Even if the presence of stretch reflexes in the human tongue has not yet been demonstrated (Neilson, Andrews, Guitar, and Quinn 1979), some neurophysiological data support the idea of a control model based on the EP Hypothesis. Indeed, a great density of muscle spindles has been found in parts of the tongue that require fine adjustment in the production of complex articulations (Cooper 1953; Walker and Rajagopal 1959). As spindles can convey proprioceptive feedback, this could support Feldman's theory.

2. METHODS

2.1. Model overview

The proposed model of the human tongue is based on the previous work of Perkell (1976). The geometry of the mid-sagittal section of the tongue is represented by a lattice of 16 mass-bearing points, or nodes, connected to each other, as well as to the mandible, the hyoid bone and the upper skull, by elastic elements (see Fig. 4). The directions of those elastic elements reflect the arrangement of the main tongue muscle fibers, namely the styloglossus, the genioglossus, the hyoglossus, the verticalis and the longitudinalis. The main differences between Perkell's model and ours are the following:

- The force generating mechanism is based on the λ model in our case.
- We did not include the palatoglossus and the pharyngeal constrictors because, in spite of the fact that they are attached to the tongue, they contribute mainly to move other tissues (velum and pharyngeal walls, respectively).
- We did not consider the mylohyoid, because it is not a tongue muscle. Actually, it is attached to

the hyoid bone and the jaw, contributing indirectly to tongue lowering/rising.

- We changed the geometry in order to adapt the model to the anatomy of the speaker with whom the cineradiographic database was obtained (Bothorel, Simon, Wioland, and Zerling 1986).

The bony parts (i.e. the mandible and the hyoid bone) can move in both horizontal and vertical directions (no rotation is allowed) and, as we are interested only on the tongue musculature, we assume that their positions are imposed to the model. Soft tissues are accounted for as muscular hydrostats (Kier and Smith 1985; Kier and Smith 1989), by assuming that variations of mid-sagittal tongue area is negligible, such that the area of each region of the lattice is preserved. This is true only as a first approximation, because only volume is expected to be conserved.

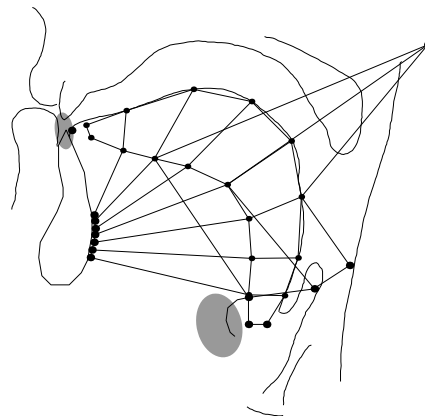


Figure 2: Schematics of the model superposed to a sagittal tracing of an X-ray frame. The flesh points are represented as smaller filled circles and the origin points for the tongue muscles are represented as bigger circles. The jaw and hyoid are represented schematically and the 3 std ellipses of dispersion for their positions are shown in gray.

2.2. Fitting procedure

A fitting procedure has been applied to a corpus of 519 mid-sagittal X-ray images (therefore called frames) of a French speaker uttering ten sentences which are phonetically equilibrated (Bothorel, Simon, Wioland, and Zerling 1986). We assume that this database is representative of all possible configurations of the tongue. Tongue profiles were obtained by hand from each 20-ms frame.

As the problem of finding flesh-point position from the tongue profile is ill-posed, it was solved by means of a regularization procedure based on both (i) an intra-frame, or spatial constraint (namely the mid-sagittal area conservation); and (ii) an inter-frame,

or temporal constraint by requiring that the change in length of the elastic segments from one frame to the next should be minimum. The goal was to fit the tongue model contour to the tongue profile by estimating the “best” placement of the flesh points. The jaw and hyoid were fixed for each frame, its positions being directly extracted from the X-ray image.

The fitting procedure yields a set of lengths for the elastic segments in each frame, and according to the principle of anatomical correspondence they will be taken as the λ commands for the associated segments. The algorithm of the fitting procedure is shown in Fig. 3.

```

for each frame do
  take the previous fitted flesh-point positions as a starting
  configuration

  change the origin positions for the muscles attached to
  the hyoid bone and the jaw

  set the position of the flesh-points corresponding to the
  superior part of the tongue to the closest points on
  the new tongue profile (spline approximation)

  compute the position of the flesh-points inside the tongue
  and on the superior part (they can slide along the
  tongue profile) satisfying (i) and (ii)
end for

```

Figure 3: Algorithm for the fitting procedure

3. RESULTS

3.1. The concept of control variables

From the mechanical point of view, the human tongue has an infinite number degrees of freedom, but tongue movements observed in normal speech appear to satisfy some kind of functional constraint that is not merely anatomical, but must be somehow related to the nature of the task. This has led to the hypothesis that vocal tract shape is determined by a small number of independently controlled components or *articulators*, whose exact nature has been long debated (Maeda 1990).

In terms of the λ -model, the control variables are indeed a set of ‘articulators’ that, different from parametric or geometrical models, have a very precise biological substrate at the control level. Therefore, the investigation of number and nature of the control variables of the tongue can provide a set of elementary gestures that have a cognitive significance, and may suggest how vocal tract shapes might be internally represented by the CNS.

3.2. Principal Component Analysis of muscle commands

It should be noted that the problem of identifying the CVs is still ill-posed, because there are infinite ways to define a mapping $\tilde{\mathcal{L}}(\vec{u})$ such that $\vec{\lambda} = \vec{l} = \tilde{\mathcal{L}}(\vec{u})$,

but it is reasonable to assume that tongue configurations are represented by the CNS so that the resulting mapping between CVs and the λ s is as ‘simple’ as possible; in particular, we will hypothesize a *linear* mapping between the CV space and the λ space, i.e.: $\vec{\lambda} \approx L \cdot \vec{u} + \vec{l}_0$. The best linear approximation of the $\vec{\lambda}$ is provided by a Principal Component Analysis (PCA), corresponding to setting the columns of L to be the eigenvectors of the covariance matrix of the set of muscle lengths \vec{l} , that can be estimated from the experimental data.

Before performing the PCA in the raw λ data, we did a linear regression in order to extract the contribution of the horizontal and vertical positions of both the hyoid and the jaw onto the λ s. This means that jaw and hyoid movements are expressed in terms of tongue muscle synergies in our model. Explicit changes of position of the bony parts reflect in changes in the muscle commands.

The PCA is then applied to the residual of the λ data after the extraction of jaw and hyoid contribution. The percentage of explanation of the variance of the whole data for each component is shown in table 1. We can see that the jaw and hyoid commands are responsible for 32.2% of the variance at the level of muscle commands. Of the 67.8% of residual variance, the four first principal components of the tongue are responsible for 83% of explanation. Only 11% of variance of the fitted λ data is not explained by the 8 components.

component	total	residual
Jaw 1 st	32.2% (32.2%)	-
Jaw 2 nd		
Hyoid 1 st		
Hyoid 2 nd		
Tongue 1 st	26.8% (59 %)	39.5% (39.5%)
Tongue 2 nd	18.6% (77.6%)	27.5% (67 %)
Tongue 3 rd	8 % (85.6%)	11.8% (78.8%)
Tongue 4 th	3.4% (89 %)	5.0% (83.8%)

Table 1: Percentage of variance explanation for each component from the PCA. (Percentages between parentheses stand for cumulative sums)

3.3. Effect of PCA commands on the tongue shape

In order to assess how the components obtained by the PCA affect the tongue shape we did simulations with the model including the muscle model. The principle of our simulation is very close to the *Passive Motion Paradigm* introduced by Morasso and Sanguineti (1991). Indeed, we applied the lambda commands to the model and by minimization of the potential elastic energy under the constraint of conservation of sagittal area, we find the equilibrium position. The results of the application of each component is shown in Fig. 4.

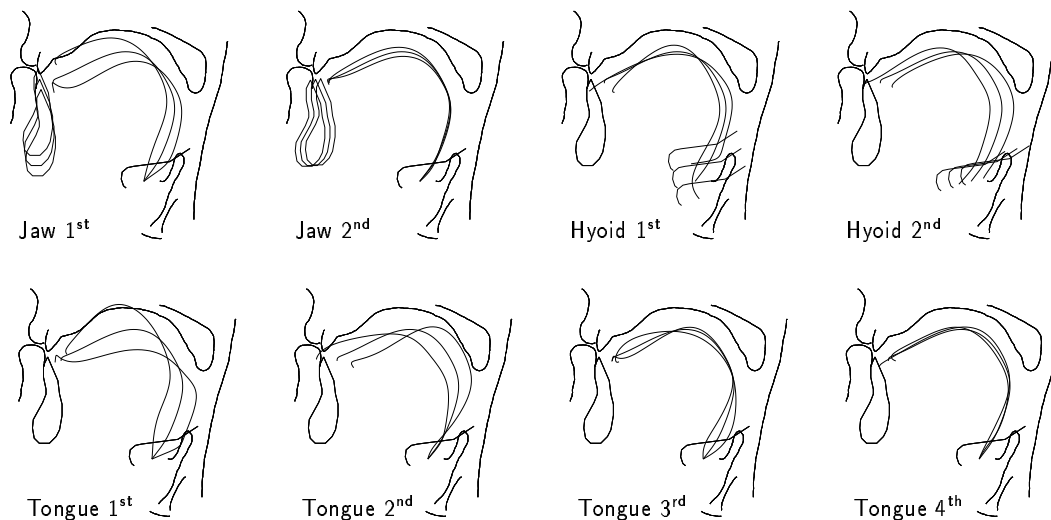


Figure 4: Effect of the commands on the tongue shape for -3 , 0 , and $+3$ standard deviations.

4. DISCUSSION

From the results shown in Fig. 4 we concluded that it is possible to associate each of the components obtained from the PCA with the main *articulators* used to describe tongue motion, namely *tongue dorsum* (arched/flattened), *tongue body* (front/back), and *tongue tip* (raised/lowered-bunched and forward/backward). The effects of our PCA commands are quite close to the effects of the commands obtained by Maeda (1988) using a statistical analysis of the tongue profile. This result suggests that (i) the main functional degrees of freedom may emerge from how the system is structured at the biomechanical level and (ii) that the CNS can use a simple strategy (invariant changes of λ s) in order to achieve movement along those degrees of freedom. We believe that the information on length of segments obtained through proprioceptive feedback is crucial for planning movements of the tongue, supporting thus the validity of the λ model as a motor control principle.

As a suggestion for further work, it would be interesting to group some elastic elements of the model, which are in series, and apply the PCA analysis on the transformed data. The rationale for doing this is that tongue muscles fibers can run through several elastic element of the model (e.g. genioglossi and longitudinalis). We should then compare the results obtained after grouping with the results presented in this paper.

A drawback of our study the fact that we perform the fitting for each frame as if the tongue position was static. In future work one should consider the dynamical effects of moving the tongue tissue, as well as gravity and palate reaction forces.

Finally we should stress the point that the λ model is highly tied to our work, even if the fitting proce-

dure was purely geometrical, no muscle model being involved. Indeed, we posited that the controlled variables are the length of the elastic elements (or muscle fibers, in the case they are grouped), and we explicitly relate tongue configuration to those muscle length, which are homogeneous to the λ s.

REFERENCES

- Bothorel, A., P. Simon, F. Wioland, and J.-P. Zerling (1986). *Cinéradiographie des voyelles et consonnes du français*. Strasbourg: Publications de l'Institut de Phonétique.
- Cooper, S. (1953). Muscle spindles in the intrinsic muscles of the human tongue. *J. of Physiology* 122, 193–202.
- Feldman, A. G. (1986). Once more on the equilibrium-point hypothesis (λ model) for motor control. *Journal of Motor Behavior* 18, 17–54.
- Feldman, A. G. (1995). Positional frames of reference in motor control: Origin and use. *Brain and Behavioral Sciences*, (in press).
- Kier, W. and K. Smith (1985). Tongues, tentacles and trunks: the biomechanics of movement in muscular hydrostats. *Zool. J. Linn. Soc.* 83, 307–324.
- Kier, W. and K. Smith (1989). Trunks, tongues and tentacles: Moving with skeletons of muscle. *Am. Sci.* 77, 29–35.
- Laboissière, R., D. J. Ostry, and A. G. Feldman (1995). The control of multi-muscle systems: Human jaw and hyoid movements. (submitted to Biological Cybernetics).
- Maeda, S. (1988). Improved articulatory model. *J. Acoust. Soc. Am.* 81(S1), S146.
- Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W. J. Hardcastle and A. Marschal (Eds.), *Speech Production and Speech Modelling*. Kluwer.
- Morasso, P. and V. Sanguineti (1991). Neurocomputing concepts in motor control. In J. Paillard (Ed.), *Brain and Space*, Chapter 21, pp. 404–432. Oxford.
- Neilson, P., G. Andrews, B. Guitar, and P. Quinn (1979). Tonic stretch reflexes in lip, tongue and jaw muscles. *Brain Research* 178, 311–327.
- Perkell, J. (1976). *A Physiologically-Oriented Model of Tongue Activity in Speech Production*. Ph. D. thesis, MIT.
- Walker, L. B. and M. D. Rajagopal (1959). Neuromuscular spindles in the human tongue. *Anatomical Record* 133, 438.